

# Hörmander's theorem for rough differential equations on manifolds

Thomas Cass\*

Martin P. Weidner†

Imperial College London

## Abstract

We introduce a new definition for solutions  $Y$  to rough differential equations (RDEs) of the form  $dY_t = V(Y_t) d\mathbf{X}_t, Y_0 = y_0$ . By using the Grossman-Larson Hopf algebra on labelled rooted trees, we prove equivalence with the classical definition of a solution advanced by Davie [12] when the state space  $E$  for  $Y$  is a finite-dimensional vector space. The notion of solution we propose, however, works when  $E$  is any smooth manifold  $\mathcal{M}$  and is therefore equally well-suited for use as an intrinsic definition of an  $\mathcal{M}$ -valued RDE solution. This enables us to prove existence, uniqueness and coordinate-invariant theorems for RDEs on  $\mathcal{M}$  bypassing the need to define a rough path on  $\mathcal{M}$ . Using this framework, we generalise result of Cass, Hairer, Litter and Tindel [9] proving the smoothness of the density of  $\mathcal{M}$ -valued RDEs driven by non-degenerate Gaussian rough paths under Hörmander's bracket condition. In doing so, we reinterpret some of the foundational results of the Malliavin calculus to make them appropriate to the study  $\mathcal{M}$ -valued Wiener functionals.

## 1 Introduction

The foundational work of Malliavin [29] gives a probabilistic scheme for proving the existence of a smooth density for the law of a hypoelliptic diffusion in  $\mathbf{R}^n$ . In the last decade, a great deal of work has been done to extend this work beyond the diffusion setting and apply it stochastic differential equations (SDEs) of the form

$$dY_t = V_0(Y_t)dt + V(Y_t) dX_t, \quad Y_0 = y_0, \quad (1)$$

where  $X$  is, typically, a continuous Gaussian process,  $V_0$  is a smooth vector field and  $V$  denotes a collection of smooth vector fields  $V = \{V_1, \dots, V_d\}$  on the state space. This

---

\*supported by EPSRC grant EP/M00516X/1

†funded by a mini-DTC scholarship from the Department of Mathematics, Imperial College London

work has been given impetus by advances in Lyons' rough path theory which provides a framework within which to interpret and manipulate such equations. The reader can consult [3, 4, 7–9, 23] for an impression of the scope of the activity in this area. At the same time there has been interest in studying rough evolutions on manifolds, see e.g. [2, 6, 10, 14]. In this paper we unify the achievements of these two strands of work, relaxing the assumption that  $Y$  is  $\mathbf{R}^n$ -valued and assuming instead that it evolves in some smooth manifold  $\mathcal{M}$ . We show how this solution can be studied and provide a natural set of conditions, analagous to those in Hörmander's theorem, under which its law has a smooth density on  $\mathcal{M}$ .

To describe the contribution of this paper, we first recall some basic principles about stochastic analysis on manifolds. In this setting one frequently proceeds by analogy with better-understood concepts in theory of vector space valued processes, see e.g. [13, 25]. As an example, consider a finite dimensional vector space  $E$  and an  $E$ -valued stochastic process  $X = (X_t)_{t \in [0, T]}$ . It is well-known that  $X$  is a semimartingale if it can be written as the sum of an (adapted) finite-variation process and a local martingale. This property can, however, also be characterised by requiring that  $(f(X_t))_{t \in [0, T]}$  is a real-valued semimartingale for every smooth real-valued function  $f$  on  $E$ . These definitions are equivalent, but the latter has two advantages over the former; first, it demands nothing of the vector space structure on  $E$ , and second it relies on no prior notion of an  $E$ -valued local martingale. Both these features make it well-suited to be the definition of a semimartingale on a smooth manifold  $\mathcal{M}$  which, while not being a linear space, nevertheless has a well-defined smooth structure. The same idea can be used to define solutions to SDEs on  $\mathcal{M}$ . Inspired by this approach, a natural definition suggests itself for giving meaning to the  $\mathcal{M}$ -valued solution to the rough differential equation (RDE)

$$dY_t = V(Y_t) d\mathbf{X}_t, \quad Y_0 = y_0, \quad (2)$$

where  $\mathbf{X}$  denotes a geometric rough path on  $\mathbb{R}^d$ . This definition is that  $Y$  should solve (2) if for all  $f \in C_c^\infty(\mathcal{M})$  we have

$$f(Y_t) - f(Y_s) \simeq \sum_{k=1}^{\lfloor p \rfloor} \sum_{i_1, \dots, i_k=1}^d V_{i_1} \cdots V_{i_k} f(Y_s) \mathbf{X}_{s,t}^{k; i_1, \dots, i_k} \quad (3)$$

where  $\{\mathbf{X}_{s,t}^{k; i_1, \dots, i_k}\}$  is the set of coordinate iterated integrals of  $\mathbf{X}_{s,t}$  of order  $k$  (see [28]). The meaning of  $\simeq$  is made precise later on in the text, but is designed to capture the fact that the right and left hand sides are the same up to terms which can be neglected. This route also allows one to bypass saying exactly what is meant by a rough path (or a controlled rough path) on  $\mathcal{M}$ , since the notion of solution is formulated in terms of well-understood objects, i.e. continuous paths, vector fields and smooth functions, which are intrinsically defined on any abstract smooth manifold

In the case where  $\mathcal{M} = \mathbf{R}^n$  or where  $\mathcal{M}$  is an embedded submanifold of  $\mathbf{R}^n$ , the question immediately arises whether (3) is consistent with the classical definitions for RDE

solutions. There are several such definitions, all of which are essentially the same but which are differently formulated; choosing the right definition to work with depends on the problem at hand and needs careful judgement. We summarise these briefly here to illustrate for context for (3) referring the reader to [15] for a more detailed discussion. Chronologically the first definition is due to Lyons [27] where a solution is a full rough path, and is identified as a fixed point of a map involving the rough integral. Davie [12], by contrast, proposes a definition under involving only the path (not the full rough path) under which  $Y$  is required to be locally well-approximated by its degree- $[p]$  Euler approximation, i.e.

$$Y_t - Y_s \simeq \sum_{k=1}^{\lfloor p \rfloor} \sum_{i_1, \dots, i_k=1}^d V_{i_1} \cdots V_{i_k} \text{id}(Y_s) \mathbf{X}_{s,t}^{k; i_1, \dots, i_k}. \quad (4)$$

In a similar vein, [20, 22] have recently characterised solutions in terms of controlled rough paths. For the purposes of this paper it is easier to work with Davie's criterion and prove equivalence with (3) when  $\mathcal{M} = \mathbf{R}^n$ .

The proof of the equivalence of (3) and (4) is more involved than it might appear at first glance. It is of course clear that (3) implies statement (4), but to prove the converse is a more delicate matter. At a basic level this difficulty arises since a general smooth function  $f$  can have non-zero derivatives of any order, while the derivatives of the identity function of degree two and higher vanish identically. The summands in (3) therefore typically involve many more terms than the corresponding ones in (4). What is needed is a systematic way of describing the terms in (4) which involve derivatives of  $f$  of a fixed degree. The algebraic structure which does this is exactly captured by the Grossman-Larson Hopf algebra on labelled rooted non-planar trees. The product in this Hopf algebra reproduces the product rule for the composition of vector fields, and it contains the shuffle Hopf algebra over the tensor algebra as a Hopf subalgebra. After a brief discussion of the background, we prove the equivalence of our new characterisation of RDE solutions with Davie's definition in Theorem 3.15. A by-product of this result is a change-of-variable formula which describes how RDEs in open subsets of  $\mathbf{R}^n$  transform under diffeomorphisms. The proof is purely combinatorial and it does not use any approximation arguments, which is why it also works for RDEs driven by branched rough paths.

From Section 4 onwards we work exclusively with manifold-valued paths and processes and develop the ground for our main result. For readers' convenience we first recall the notion of a density on  $\mathcal{M}$ , and then use the results of Section 3 to prove the existence, uniqueness and coordinate-invariance for  $\mathcal{M}$ -valued RDE solutions. In Section 6 we develop the tools of Malliavin calculus. These mostly follow the classical definitions for Wiener functionals, but some important adaptations need to be made. First we need to identify the class of Malliavin-smooth random variables. This is usually defined as the  $L^p$ -completion of the set of cylindrical random variables but, since there is no analogue of  $L^p$ -spaces on a general smooth manifold, this definition needs to be refashioned. To work around this, we define an  $\mathcal{M}$ -valued random variable to be Malliavin smooth if its composition

with any smooth compactly supported function on  $\mathcal{M}$  is Malliavin smooth in the classical sense. This allows the Malliavin derivative of a Wiener functional  $F$  to be identified as a bounded linear map from the Cameron-Martin space  $\mathcal{H}$  to the tangent space  $T_{F(\omega)}\mathcal{M}$ . The question again arises whether this notion of smoothness coincides with the classical one e.g. when  $\mathcal{M} = \mathbf{R}$ . In this case, the answer is negative and we provide an explicit example to show that our new class is strictly larger. The reason for this difference again is that no global integrability restriction is imposed on the Wiener functional; smoothness is a property determined by testing against compactly supported functions and hence is local in nature. It is an interesting consequence of our proof that integrability of  $F$  does not seem needed. This seems to have gone unnoticed in the literature and which may have implications elsewhere.

As usual, smoothness of the density rests upon the strong non-degeneracy of the Malliavin covariance matrix. The covariance matrix for our notion of derivative requires that we work with a Riemannian metric. Importantly, however, the criterion for density-smoothness we derive is independent of the choice of metric; we thus rely only on its existence, a fact that is guaranteed for any smooth manifold. In the final section we combine this analysis to study the smoothness of the density for  $Y_t$  solving (2) when the driving noise is the rough path lift of continuous zero-mean Gaussian process  $X$  to a geometric rough path. In  $\mathbf{R}^n$  a natural set of non-degeneracy conditions on  $X$  for this to hold has been identified in [9]. We finish the paper by generalising this result to  $\mathcal{M}$  under the same conditions on  $X$  and assuming that  $\{V_1, \dots, V_d\}$  satisfy Hörmander's condition:

$$\text{span} \{V_i(y_0), [V_i, V_j](y_0), [V_i, [V_j, V_k]](y_0), \dots : i, j, k, \dots = 1, \dots, d\} = T_{y_0}\mathcal{M}. \quad (5)$$

A similar result for diffusion processes on a Riemannian manifold has been derived by Taniguchi [35] under a version of (5) which is uniform on  $\mathcal{M}$ . Driver [13] has also obtained analogous results, again for Riemannian manifolds. Our approach is inspired by, and uses techniques from, both of these papers.

## 2 Rough differential equations and the tensor algebra

We assume that the reader is familiar with the theory of rough paths and rough differential equations and highlight those aspects of the theory that are especially important for the present work. For details on the algebraic background see [28, 33].

**Definition 2.1** For a given vector space  $E$  we denote by

$$T((E)) := \{(a^0, a^1, a^2, \dots) : a^k \in E^{\otimes k} \text{ for } k \in \mathbf{N}_0\}$$

the algebra of *formal tensor series* over  $E$ , by

$$T(E) := \{a \in T((E)) : a^k = 0 \text{ for almost all } k \in \mathbf{N}_0\}$$

the *tensor algebra* over  $E$  and by

$$T^{(n)}(E) := \{(a^0, a^1, \dots, a^n) : a^k \in E^{\otimes k} \text{ for } k \in \{0, 1, \dots, n\}\}$$

the *truncated tensor algebra* over  $E$  for any  $n \in \mathbf{N}$ .

Furthermore we denote by  $\tilde{T}((E))$ ,  $\tilde{T}(E)$  and  $\tilde{T}^{(n)}(E)$  the subsets of  $T((E))$ ,  $T(E)$  and  $T^{(n)}(E)$  respectively whose elements satisfy  $a^0 = 1$ .

Note that  $T^{(n)}(E)$  is a subspace of  $T(E)$  and  $T((E))$  but not a subalgebra. Its multiplication is given by the usual tensor product where all terms of order greater than  $n$  are discarded. Furthermore  $T(E)$  can equivalently be defined as the free algebra that is generated by a basis of  $E$ .

For us the underlying vector space will always be  $\mathbf{R}^d$ . It is well known that  $T((\mathbf{R}^d))$  can be turned into a bialgebra (in fact: a Hopf algebra) by introducing the *deshuffle coproduct*  $\Delta_{\sqcup}$ . Instead of giving its definition, we describe its dual, namely the *shuffle product*  $\sqcup := \Delta_{\sqcup}^*$  on  $T((\mathbf{R}^d))^* \simeq T(\mathbf{R}^{d*})$ . For  $e_1^*, \dots, e_k^*, \dots, e_l^* \in \mathbf{R}^{d*}$  it is given by

$$e_1^* \otimes \dots \otimes e_k^* \sqcup e_{k+1}^* \otimes \dots \otimes e_l^* := \sum_{\sigma \in \text{Sh}(k, l-k)} e_{\sigma^{-1}(1)}^* \otimes \dots \otimes e_{\sigma^{-1}(l)}^*,$$

where the sum ranges over all permutations of  $l$  elements that preserve the order of the first  $k$  and the last  $l-k$  elements. Since  $\sqcup$  is the dual of  $\Delta_{\sqcup}$ , we have that an element  $x \in \tilde{T}((\mathbf{R}^d))$  is *group-like*, i.e. it satisfies  $\Delta_{\sqcup}(x) = x \otimes x$ , if and only if  $e^* \sqcup f^*(x) = e^*(x)f^*(x)$  holds for all  $e^*, f^* \in T(\mathbf{R}^{d*})$ .

Before we can give the definition of a rough path we need the notion of a *control*, which is a non-negative continuous function  $\omega$  defined on the set  $\{(s, t) \in [0, T]^2 : s < t\}$  with the property that  $\omega(s, t) + \omega(t, u) \leq \omega(s, u)$  and  $\omega(s, s) = 0$  for all  $s < t < u$ .

**Definition 2.2** Let  $p \geq 1$  and let  $\omega$  be a control. A continuous map  $\mathbf{X} : \{(s, t) \in [0, T]^2 : s < t\} \rightarrow T^{(\lfloor p \rfloor)}(\mathbf{R}^d)$  is called a *weakly geometric  $p$ -rough path controlled by  $\omega$*  if it satisfies the following properties.

- (i) We have  $\mathbf{X}_{s,t} \otimes \mathbf{X}_{t,u} = \mathbf{X}_{s,u}$  for all  $s < u < t$ .
- (ii)  $\mathbf{X}_{s,t}$  is the truncation of a group-like element of  $T((\mathbf{R}^d))$  for all  $s < t$ .
- (iii)  $|\mathbf{X}_{s,t}^k| \leq \omega(s, t)^{\frac{k}{p}}$  holds for all  $s < t$  and all  $k \in \{1, \dots, \lfloor p \rfloor\}$ .

**Remark 2.3** Property (ii) implies that for  $e^* \in (\mathbf{R}^{d*})^{\otimes k}$  and  $f^* \in (\mathbf{R}^{d*})^{\otimes l}$  with  $k+l \leq \lfloor p \rfloor$  we have  $e^* \sqcup f^*(\mathbf{X}_{s,t}) = e^*(\mathbf{X}_{s,t})f^*(\mathbf{X}_{s,t})$  for all  $s < t$ .

Now consider a linear map  $V : \mathbf{R}^d \rightarrow \Gamma(\mathbf{R}^n)$ , which takes values in the space of smooth vector fields on  $\mathbf{R}^n$ . In other words,  $V_a$  is a linear first order differential operator on  $\mathbf{R}^n$  for every  $a \in \mathbf{R}^d$ .

The space  $\text{Diff}(\mathbf{R}^n)$  of linear differential operators forms an algebra under the usual composition of operators and therefore we can extend  $V$  to an algebra morphism  $\mathcal{V}$  from the free algebra that is generated by  $\mathbf{R}^d$  (i.e. its tensor algebra  $T(\mathbf{R}^d)$ ) to  $\text{Diff}(\mathbf{R}^n)$ . We denote by  $\mathcal{V}_a$  the image of  $a \in T(\mathbf{R}^d)$  under this map.

We want to give meaning to the rough differential equation

$$dY = V(Y)d\mathbf{X}, \quad Y_0 = y_0 \in \mathbf{R}^n, \quad (6)$$

where  $\mathbf{X}$  is a weakly geometric  $p$ -rough path controlled by some control  $\omega$ .

**Definition 2.4** A continuous path  $Y : [0, T] \rightarrow \mathbf{R}^n$  is called a *solution of (6)* if there exists a function  $\theta : \mathbf{R} \rightarrow \mathbf{R}$  with  $\theta(h)/h \rightarrow 0$  for  $h \rightarrow 0$  such that

$$|Y_t - \mathcal{V}_{\mathbf{X}_{s,t}} \text{id}(Y_s)| \leq \theta(\omega(s, t)) \quad (7)$$

holds for all  $0 \leq s < t \leq T$ .

**Remark 2.5** In light of [16, Remark 10.18] it seems that there should be  $Y_t - Y_s$  instead of  $Y_t$  in the above definition. Note however that we have  $\mathcal{V}_1 \text{id}(Y_s) = Y_s$  and hence the seemingly missing term is contained in  $\mathcal{V}_{\mathbf{X}_{s,t}} \text{id}(Y_s)$ .

This definition is due to Davie [12], see also [16]. In [22] it has been shown (in the more general context of equations driven by branched rough paths) that this definition is equivalent to Gubinelli's definition which uses controlled rough paths [15, 19]. Davie's definition is perfectly suited for being generalised to RDEs on manifolds as we will see in Theorem 3.15 and its corollaries.

The following statement about existence and uniqueness of solutions is a combination of [16, Theorem 10.14] and [16, Theorem 10.26], where the results are stated in much more detail. See also [16, Remark 10.18] for the relation between the various definitions of RDE solutions.

**Theorem 2.6** *Let  $\mathbf{X}$  be a weakly geometric rough path over  $\mathbf{R}^d$  and let  $V : \mathbf{R}^d \rightarrow \Gamma(\mathbf{R}^n)$  be a linear map. Assume that the vector fields  $V_a$  and their derivatives of degree  $\lfloor p \rfloor$  or lower are bounded for every  $a \in \mathbf{R}^d$ . Then (6) has a unique solution on any interval  $[0, T]$  on which  $\mathbf{X}$  is defined.*

### 3 A change of variable formula for weakly geometric rough paths

With the notation and the assumptions of the preceding section, we consider the map

$$\Psi_{V,f,y} : T(\mathbf{R}^d) \rightarrow W : \mathbf{x} \mapsto \mathcal{V}_{\mathbf{x}} f(y)$$

for any sufficiently smooth function  $f$  defined on  $\mathbf{R}^n$  with values in a finite-dimensional vector space  $W$ . In order to understand the combinatorial properties of this map, we will split it as the composition of two maps

$$\Psi_{V,f,y} = \hat{\Psi}_{V,f,y} \circ \Phi,$$

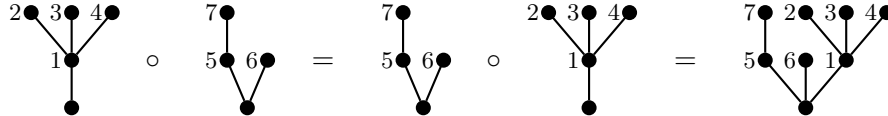
where  $\Phi$  is a universal map into the *Grossman-Larson algebra* of labelled rooted trees, to be introduced now. We begin with some notation

**Definition 3.1** We denote by

- (i)  $\mathcal{L}_d\mathcal{T}$  the set of non-planar rooted trees whose nodes (except for the root) are labelled by the set  $\{1, \dots, d\}$ ,
- (ii)  $|\mathfrak{t}|$  the number of nodes (not including the root) of  $\mathfrak{t} \in \mathcal{L}_d\mathcal{T}$ ,
- (iii)  $\|\mathfrak{t}\|$ , the number of children of the root of  $\mathfrak{t} \in \mathcal{L}_d\mathcal{T}$ ,
- (iv)  $\mathcal{L}_d\mathcal{T}^k$  the set  $\{\mathfrak{t} \in \mathcal{L}_d\mathcal{T} : |\mathfrak{t}| = k\}$ ,
- (v)  $\mathcal{L}_d\mathcal{T}_n$  the set  $\{\mathfrak{t} \in \mathcal{L}_d\mathcal{T} : \|\mathfrak{t}\| = n\}$ ,
- (vi)  $\mathcal{L}_d\mathcal{T}_n^k$  the set  $\mathcal{L}_d\mathcal{T}_n \cap \mathcal{L}_d\mathcal{T}^k$ .

A family of trees is called a *forest*. There are two useful maps  $B_+$  and  $B_-$  that map forests to trees and vice versa. The map  $B_+$  maps a forest  $\{\mathfrak{t}_1, \dots, \mathfrak{t}_n\}$  to the tree which is obtained by joining all of the roots of  $\mathfrak{t}_1, \dots, \mathfrak{t}_n$  to one new root.  $B_-$  is the inverse of this map, i.e. it maps a tree to the forest which is obtained by removing the root and the adjacent edges. Note that, technically, for  $\mathfrak{t} \in \mathcal{L}_d\mathcal{T}$  the trees that are contained in  $B_-(\mathfrak{t})$  are not elements of  $\mathcal{L}_d\mathcal{T}$  since their roots are labelled. Nevertheless we can define the *merging product*  $\circ : \mathcal{L}_d\mathcal{T} \times \mathcal{L}_d\mathcal{T} \rightarrow \mathcal{L}_d\mathcal{T}$  by  $\mathfrak{t}_1 \circ \mathfrak{t}_2 := B_+(B_-(\mathfrak{t}_1), B_-(\mathfrak{t}_2))$ , i.e. the merging product identifies the roots of the trees  $\mathfrak{t}_1$  and  $\mathfrak{t}_2$ .

**Example 3.2**

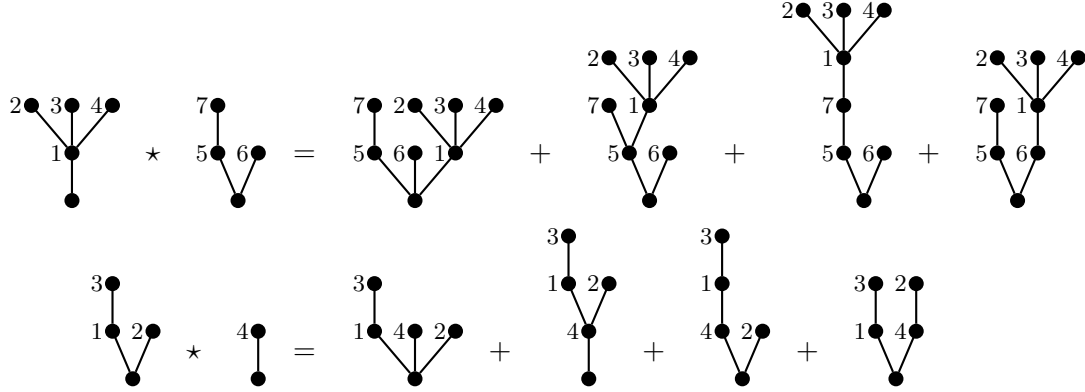


In this example and in the ones below we have chosen a different label for every node in order to make the operations more obvious. However, different nodes of one tree can of course carry the same label.

Now we consider the vector space  $\mathbf{R}\{\mathcal{L}_d\mathcal{T}\}$  that is spanned by  $\mathcal{L}_d\mathcal{T}$  and equip it with a Hopf algebra structure as follows.

Let  $\mathfrak{t}_1 \in \mathcal{L}_d\mathcal{T}_k$  and  $\mathfrak{t}_2 \in \mathcal{L}_d\mathcal{T}_n$  be two trees. There are  $(n+1)^k$  possibilities to graft all of the trees of the forest  $B_-(\mathfrak{t}_1)$  to the nodes of  $\mathfrak{t}_2$  by new edges and all of these trees are elements of  $\mathcal{L}_d\mathcal{T}$ . Define the product  $\mathfrak{t}_1 \star \mathfrak{t}_2$  as the sum over all the trees that are obtained in this way.

### Example 3.3

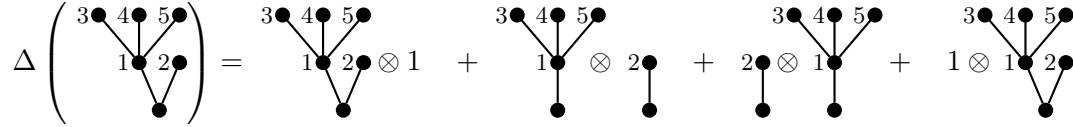


The coproduct  $\Delta(\mathbf{t})$  of a tree  $\mathbf{t} \in \mathcal{L}_d\mathcal{T}_k$  is given by

$$\Delta(\mathbf{t}) = \sum_{I \subset \{1, \dots, k\}} B_+(B_-(\mathbf{t})_I) \otimes B_+(B_-(\mathbf{t})_{I^c}),$$

where  $I^c := \{1, \dots, k\} \setminus I$  and  $B_-(\mathbf{t})_I$  is the subforest of  $B_-(\mathbf{t})$  that corresponds to  $I$ . By this we mean that we fix an ordering of the set  $B_-(\mathbf{t})$  and then we denote by  $B_-(\mathbf{t})_I$  the subset which contains the  $i$ -th element of  $B_-(\mathbf{t})$  if and only if  $i \in I$ .

### Example 3.4



One can easily see that the elements in  $\mathcal{L}_d\mathcal{T}_1$  are primitive, i.e. they satisfy

$$\Delta(\mathbf{t}) = \mathbf{t} \otimes 1 + 1 \otimes \mathbf{t},$$

where we write 1 for the tree that consists only of the root.

Finally we define the counit  $\epsilon(\mathbf{t})$  to be 1 if  $\mathbf{t}$  is the tree that only consists of the root and 0 else and we denote by  $\eta$  the map which takes  $1 \in \mathbf{R}$  to the tree that consists only of the root.

**Proposition 3.5** ([18, Theorem 3.2])  $\mathcal{H}_{GL} := (\bigoplus_{k=0}^{\infty} \mathbf{R}\{\mathcal{L}_d\mathcal{T}^k\}, \star, \eta, \Delta, \epsilon)$  is a cocommutative, graded connected bialgebra and therefore it is also a Hopf algebra. The grading is given by  $|\cdot|$ .

Consider the map  $\mathbf{R}^d \rightarrow \mathcal{H}_{GL}$  which maps the basis vector  $e_i$  to the tree which consists of the root and one node, where the node carries the label  $i$ . This map extends uniquely to our desired universal algebra morphism  $\Phi : T(\mathbf{R}^d) \rightarrow \mathcal{H}_{GL}$ .



### Example 3.6

$$\Phi(e_1 \otimes e_2 \otimes e_3) = \begin{array}{c} 1 \\ \bullet \\ | \\ \bullet \end{array} \star \begin{array}{c} 2 \\ \bullet \\ | \\ \bullet \end{array} \star \begin{array}{c} 3 \\ \bullet \\ | \\ \bullet \end{array} = \begin{array}{c} 1 \\ \bullet \\ | \\ \bullet \\ \diagup \quad \diagdown \\ 2 \bullet \quad 3 \bullet \end{array} + \begin{array}{c} 1 \\ \bullet \\ | \\ \bullet \\ \diagup \quad \diagdown \\ 2 \bullet \quad 3 \bullet \end{array} + \begin{array}{c} 1 \\ \bullet \\ | \\ \bullet \\ \diagup \quad \diagdown \\ 1 \bullet \quad 2 \bullet \quad 3 \bullet \end{array} + \begin{array}{c} 1 \\ \bullet \\ | \\ 2 \bullet \\ | \\ 3 \bullet \end{array} + \begin{array}{c} 1 \bullet \quad 2 \bullet \\ \diagdown \quad \diagup \\ 3 \bullet \\ | \\ \bullet \end{array} + \begin{array}{c} 2 \bullet \\ | \\ 1 \bullet \quad 3 \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array}$$

**Proposition 3.7** *The algebra morphism  $\Phi$  is a Hopf algebra morphism.*

*Proof.* We have to show that  $\Delta \circ \Phi = (\Phi \otimes \Phi) \circ \Delta_{\sqcup}$ . It suffices to prove that this equality holds on a generator of  $T(\mathbf{R}^d)$ . Since  $\mathbf{R}^d$  generates  $T(\mathbf{R}^d)$  and all elements of  $\mathbf{R}^d$  are primitive with respect to  $\Delta_{\sqcup}$ , we only have to show that  $\Phi(v)$  is primitive with respect to  $\Delta$  for every  $v \in \mathbf{R}^d$ . We have already noted above that elements of  $\mathcal{L}_d \mathcal{T}_1$  are primitive, hence the conclusion follows from the observation that  $\Phi(v) \in \mathbf{R}\{\mathcal{L}_d \mathcal{T}_1\}$  for all  $v \in \mathbf{R}^d$ . ■

**Remark 3.8** The map  $\Phi$  is used in [22] to show that weakly geometric rough paths can be viewed as branched rough paths in a canonical way.

Let us now see how the map  $\hat{\Psi}_{V,f,y}$  is constructed. A tree  $\mathfrak{t} \in \mathcal{L}_d \mathcal{T}$  is mapped to a certain product of derivatives of  $f$  and  $V$  evaluated at  $y$  in the following way. The root corresponds to  $f$  and a node labelled by  $i \in \{1, \dots, d\}$  corresponds to the vector field  $V_{e_i}$ . The number of children of a node indicates how often the corresponding function or vector field has to be differentiated. These derivatives are then evaluated at  $y$  which leads to symmetric multilinear maps, into which we plug the children.

### Example 3.9

$$\hat{\Psi}_{V,f,y} \left( \begin{array}{c} 3 \bullet \quad 4 \bullet \quad 5 \bullet \\ \diagdown \quad \diagup \quad \diagup \\ 1 \bullet \quad 2 \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array} \right) = f''(y)[V_1'''(y)[V_3(y), V_4(y), V_5(y)], V_2(y)]$$

One can now easily see that we have indeed

$$\Psi_{V,f,y} = \hat{\Psi}_{V,f,y} \circ \Phi. \quad (8)$$

The key to this observation is the fact that the product  $\star$  mimics exactly the product rule for the composition of two linear differential operators.

There is a particular choice of  $f$  which will become important. For any positive integer  $k$  we define the map

$$\text{id}^k : \mathbf{R}^n \rightarrow (\mathbf{R}^n)^{\otimes k} : x \mapsto x^{\otimes k}.$$

The maps  $\hat{\Psi}_{V,\text{id}^k,y}$  have two important properties. First they satisfy

$$\hat{\Psi}_{V,f,y}(\mathfrak{t}) = \frac{1}{k!} f^{(k)}(y)[\hat{\Psi}_{V,\text{id}^k,y}(\mathfrak{t})] \quad (9)$$

for all  $\mathbf{t} \in \mathcal{L}_d \mathcal{T}_k$  and all sufficiently differentiable functions  $f$ . Second they are compatible with the operation of merging two trees in the sense that for two trees  $\mathbf{t}_1 \in \mathcal{L}_d \mathcal{T}_k$  and  $\mathbf{t}_2 \in \mathcal{L}_d \mathcal{T}_l$  we have

$$\hat{\Psi}_{V, \text{id}^{k+l}, y}(\mathbf{t}_1 \circ \mathbf{t}_2) = \frac{(k+l)!}{k!l!} \text{Sym} \left( \hat{\Psi}_{V, \text{id}^k, y}(\mathbf{t}_1) \otimes \hat{\Psi}_{V, \text{id}^l, y}(\mathbf{t}_2) \right). \quad (10)$$

We state the following lemma because of its importance for what follows, even though the statement is completely obvious.

**Lemma 3.10** *For every tree  $\mathbf{t} \in \mathcal{L}_d \mathcal{T}$  there exist unique trees  $\mathbf{t}_1, \dots, \mathbf{t}_k \in \mathcal{L}_d \mathcal{T}_1$  such that  $\mathbf{t} = \mathbf{t}_1 \circ \dots \circ \mathbf{t}_k$ .*

**Definition 3.11** The *graded dual* of  $\mathcal{H}_{GL}$  is given by the vector space

$$\mathcal{H}_{GL}^{gr} := \bigoplus_{k=0}^{\infty} \mathbf{R}\{\mathcal{L}_d \mathcal{T}^k\}^*$$

which is then equipped with the product  $\Delta^*$ .

The graded dual can of course be turned into a Hopf algebra but we are only interested in its algebra structure. Since  $\mathcal{H}_{GL}$  is locally finite (i.e. the subspace corresponding to a fixed degree is finite-dimensional) it is isomorphic to its graded dual as a vector space. Therefore a basis for  $\mathcal{H}_{GL}^{gr}$  is given by the linear forms of the type  $\varphi_{\mathbf{t}}$  which are defined by

$$\varphi_{\mathbf{t}}(\mathbf{s}) := \delta_{\mathbf{s}\mathbf{t}}$$

for  $\mathbf{s} \in \mathcal{L}_d \mathcal{T}$ . By definition, the product  $\varphi_{\mathbf{t}_1} \Delta^* \varphi_{\mathbf{t}_2}$  satisfies

$$(\varphi_{\mathbf{t}_1} \Delta^* \varphi_{\mathbf{t}_2})(\mathbf{s}) = \varphi_{\mathbf{t}_1} \otimes \varphi_{\mathbf{t}_2} \Delta(\mathbf{s})$$

for every  $\mathbf{s} \in \mathcal{L}_d \mathcal{T}$ . This shows immediately that  $\varphi_{\mathbf{t}_1} \Delta^* \varphi_{\mathbf{t}_2} = N \varphi_{\mathbf{t}_1 \circ \mathbf{t}_2}$  for some integer  $N$  which depends on  $\mathbf{t}_1$  and  $\mathbf{t}_2$ . In fact, we have the following result.

**Proposition 3.12** *Assume that  $\mathbf{t} \in \mathcal{L}_d \mathcal{T}$  can be written as the merging product of  $k$  trees  $\mathbf{t}_1, \dots, \mathbf{t}_k \in \mathcal{L}_d \mathcal{T}_1$  and assume that there exist integers  $1 = i_1 < i_2 < \dots < i_r \leq k$  such that*

$$\mathbf{t}_{i_1} = \dots = \mathbf{t}_{i_2-1}, \mathbf{t}_{i_2} = \dots = \mathbf{t}_{i_3-1}, \dots, \mathbf{t}_{i_r} = \dots = \mathbf{t}_k$$

*holds and such that the trees  $\mathbf{t}_{i_1}, \mathbf{t}_{i_2}, \dots, \mathbf{t}_{i_r}$  are pairwise distinct. Then we have*

$$\varphi_{\mathbf{t}_1} \Delta^* \dots \Delta^* \varphi_{\mathbf{t}_k} = N(\mathbf{t}) \varphi_{\mathbf{t}},$$

*where  $N(\mathbf{t}) := (i_2 - i_1)! \dots (i_r - i_{r-1})!$ .*

*Proof.* We use the associativity of  $\Delta^*$  and assume first that  $r = 1$ , i.e. that we have  $k$  identical trees. We prove this case by induction over  $k$ . For  $k = 1$  there is nothing to show. For  $k > 1$  we have

$$\varphi_{t_1} \Delta^* \cdots \Delta^* \varphi_{t_k} = (k-1)! \cdot \varphi_{t_1 \circ \cdots \circ t_{k-1}} \Delta^* \varphi_{t_k} = (k-1)! \cdot k \cdot \varphi_{t_1 \circ \cdots \circ t_k},$$

where the last equality holds because the term  $t_1 \circ \cdots \circ t_{k-1} \otimes t_k$  appears exactly  $k$  times in the co-product of  $t_1 \circ \cdots \circ t_k$ .

If  $r$  is different from 1, then we define  $\bar{t}_1 := t_{i_1} \circ \cdots \circ t_{i_2-1}$ ,  $\bar{t}_2 := t_{i_2} \circ \cdots \circ t_{i_3-1}$ ,  $\dots$ ,  $\bar{t}_r := t_{i_r} \circ \cdots \circ t_k$ . Since we assumed that  $t_{i_1}, t_{i_2}, \dots, t_{i_r}$  are pairwise distinct, one easily sees that we have  $\varphi_{\bar{t}_1} \Delta^* \cdots \Delta^* \varphi_{\bar{t}_r} = \varphi_{\bar{t}_1} \circ \cdots \circ \varphi_{\bar{t}_r}$ , whence the claim follows.  $\blacksquare$

**Remark 3.13** Since the merging product  $\circ$  is essentially the free product of the Connes-Kreimer Hopf algebra, the preceding proposition indicates that the Grossman-Larson Hopf algebra is the dual of the Connes-Kreimer Hopf algebra (and vice versa). A detailed proof of this duality is given in [24]

An element  $a \in \mathcal{H}_{GL}$  is a linear combination of (finitely many) trees. Therefore we can define  $a|_k$  as the element of  $\mathcal{L}_d \mathcal{T}_k$  which is obtained by removing all the terms from  $a$  that are not elements of  $\mathcal{L}_d \mathcal{T}_k$ .

**Lemma 3.14** *Let  $\mathbf{x} \in T^{(n)}(\mathbf{R}^d)$  be the truncation of a group-like element of  $T(\mathbf{R}^d)$ . Then we have*

$$(\Phi(\mathbf{x})|_1)^{\circ k} = k! \Phi(\mathbf{x})|_k + \mathfrak{R}_{>n}$$

where  $\mathfrak{R}_{>n} \in \mathcal{H}_{GL}$  is a linear combination of trees with more than  $n$  nodes.

*Proof.* We show that  $\varphi_{\mathbf{t}}((\Phi(\mathbf{x})|_1)^{\circ k}) = k! \varphi_{\mathbf{t}}(\Phi(\mathbf{x}))$  holds for all  $\mathbf{t} \in \mathcal{L}_d \mathcal{T}_k^m$  and all  $m \leq n$ .

Thus let  $m \leq n$ . Consider the decomposition of  $\mathbf{t} \in \mathcal{L}_d \mathcal{T}_k^m$  into  $\mathbf{t} = t_1 \circ \cdots \circ t_k$  with  $t_1, \dots, t_k \in \mathcal{L}_d \mathcal{T}_1$  (recall Lemma 3.10). Then we have

$$\begin{aligned} \varphi_{\mathbf{t}}((\Phi(\mathbf{x})|_1)^{\circ k}) &= \frac{k!}{N(\mathbf{t})} \varphi_{t_1}(\Phi(\mathbf{x})) \cdots \varphi_{t_k}(\Phi(\mathbf{x})) \\ &= \frac{k!}{N(\mathbf{t})} \varphi_{t_1} \Delta^* \cdots \Delta^* \varphi_{t_k}(\Phi(\mathbf{x})) \\ &= k! \varphi_{\mathbf{t}}(\Phi(\mathbf{x})). \end{aligned}$$

The first equality is simply an application of the multinomial formula. The second equality holds because  $\mathbf{x}$  is the truncation of a group-like element and  $\Phi$  is a Hopf-algebra morphism (see Remark 2.3). The last equality is an application of Proposition 3.12.  $\blacksquare$

**Theorem 3.15** *Let  $\mathbf{X}$  be a weakly geometric  $p$ -rough path in  $\mathbf{R}^d$  which is controlled by a control  $\omega$  and let  $V : \mathbf{R}^n \rightarrow \mathcal{L}(\mathbf{R}^d, \mathbf{R}^n)$  be a smooth map. Then, if  $Y : [0, T] \rightarrow \mathbf{R}^n$  is a continuous path, the following are equivalent*

(i)  $Y$  solves the equation  $dY = V(Y)d\mathbf{X}$ , i.e. there exists a function  $\theta : \mathbf{R} \rightarrow \mathbf{R}$  with  $\theta(h)/h \rightarrow 0$  for  $h \rightarrow 0$  such that

$$|Y_t - \mathcal{V}_{\mathbf{X}_{s,t}} \text{id}(Y_s)| \leq \theta(\omega(s, t)) \quad (11)$$

holds for all  $0 \leq s < t \leq T$ .

(ii) For any smooth function  $f : \mathbf{R}^n \rightarrow W$  which takes values in some vector space  $W$  we have

$$|f(Y_t) - \mathcal{V}_{\mathbf{X}_{s,t}} f(Y_s)| \leq \theta_f(\omega(s, t)) \quad (12)$$

for all  $0 \leq s < t \leq T$  and a function  $\theta_f : \mathbf{R} \rightarrow \mathbf{R}$  with  $\theta_f(h)/h \rightarrow 0$  for  $h \rightarrow 0$ .

*Proof.* The implication (ii)  $\Rightarrow$  (i) is trivial. For the other direction, the idea is to split the expression  $\mathcal{V}_{\mathbf{X}_{s,t}} f(Y_s)$  into terms that we can control. We have (see explanations below)

$$\begin{aligned} & |f(Y_t) - \mathcal{V}_{\mathbf{X}_{s,t}} f(Y_s)| \\ &= \left| f(Y_t) - \hat{\Psi}_{V, f, Y_s}(\Phi(\mathbf{X}_{s,t})) \right| \end{aligned} \quad (13a)$$

$$\leq \left| \sum_{k=1}^{\lfloor p \rfloor} \frac{f^{(k)}(Y_s)}{k!} \left[ Y_{s,t}^{\otimes k} - \hat{\Psi}_{V, \text{id}^k, Y_s}(\Phi(\mathbf{X}_{s,t})|_k) \right] \right| + C_1 |Y_{s,t}|^{\lfloor p \rfloor + 1} \quad (13b)$$

$$\leq \left| \sum_{k=1}^{\lfloor p \rfloor} \frac{f^{(k)}(Y_s)}{k!} \left[ Y_{s,t}^{\otimes k} - \frac{1}{k!} \hat{\Psi}_{V, \text{id}^k, Y_s} \left( (\Phi(\mathbf{X}_{s,t})|_1)^{\otimes k} \right) \right] \right| + C_1 |Y_{s,t}|^{\lfloor p \rfloor + 1} + C_2 \omega(s, t)^{\frac{\lfloor p \rfloor + 1}{p}} \quad (13c)$$

$$= \left| \sum_{k=1}^{\lfloor p \rfloor} \frac{f^{(k)}(Y_s)}{k!} \left[ Y_{s,t}^{\otimes k} - \left( \hat{\Psi}_{V, \text{id}, Y_s}(\Phi(\mathbf{X}_{s,t})|_1) \right)^{\otimes k} \right] \right| + C_1 |Y_{s,t}|^{\lfloor p \rfloor + 1} + C_2 \omega(s, t)^{\frac{\lfloor p \rfloor + 1}{p}} \quad (13d)$$

$$\begin{aligned} &= \left| \sum_{k=1}^{\lfloor p \rfloor} \frac{f^{(k)}(Y_s)}{k!} \left[ \left( Y_{s,t} - \hat{\Psi}_{V, \text{id}, Y_s}(\Phi(\mathbf{X}_{s,t})|_1) \right) \otimes \sum_{i=0}^{k-1} Y_{s,t}^{\otimes (k-1-i)} \hat{\Psi}_{V, \text{id}, Y_s}(\Phi(\mathbf{X}_{s,t})|_1)^{\otimes i} \right] \right| \\ &\quad + C_1 |Y_{s,t}|^{\lfloor p \rfloor + 1} + C_2 \omega(s, t)^{\frac{\lfloor p \rfloor + 1}{p}} \end{aligned} \quad (13e)$$

$$\begin{aligned} &= \left| \sum_{k=1}^{\lfloor p \rfloor} \frac{f^{(k)}(Y_s)}{k!} \left[ (Y_t - \mathcal{V}_{\mathbf{X}_{s,t}} \text{id}(Y_s)) \otimes \sum_{i=0}^{k-1} Y_{s,t}^{\otimes (k-1-i)} \mathcal{V}_{\mathbf{X}_{s,t}} \text{id}(Y_s)^{\otimes i} \right] \right| \\ &\quad + C_1 |Y_{s,t}|^{\lfloor p \rfloor + 1} + C_2 \omega(s, t)^{\frac{\lfloor p \rfloor + 1}{p}}. \end{aligned} \quad (13f)$$

In this computation we have used the following arguments.

(13a) Rewrite  $\mathcal{V}_{\mathbf{X}_{s,t}} f(Y_s)$  in terms of the functions  $\hat{\Psi}$  and  $\Phi$  that we have introduced above, recall (8).

- (13b) Replace  $f$  by its Taylor approximation up to degree  $\lfloor p \rfloor$  and make use of (9).
- (13c) Apply Lemma 3.14, where the remainder  $\mathfrak{R}_{>\lfloor p \rfloor}$  leads to the term that involves  $\omega$ .
- (13d) Use (10), where we do not have to take the symmetric part because  $f^{(k)}(Y_s)$  is a symmetric bilinear form anyway.
- (13e) Use the elementary identity  $a^k - b^k = (a - b)(a^{k-1} + a^{k-2}b + \dots + ab^{k-2} + b^{k-1})$ , which holds in principle only for commuting variables  $a$  and  $b$ . Can be applied nevertheless, again because  $f^{(k)}(Y_s)$  is symmetric.
- (13f) Observe that  $\hat{\Psi}_{V, \text{id}, Y_s}$  maps any tree which is not either in  $\mathcal{L}_d \mathcal{T}_1$  or the trivial tree which consists only of the root to zero.

If we define the constant  $C_3$  by

$$C_3 := \sup_{0 \leq s < t \leq T} \sup_{a \in \mathbf{R}^n: \|a\|=1} \left| \sum_{k=1}^{\lfloor p \rfloor} \frac{f^{(k)}(Y_s)}{k!} \left[ a \otimes \sum_{i=0}^{k-1} Y_{s,t}^{\otimes(k-1-i)} \mathcal{V}_{\mathbf{X}_{s,t}} \text{id}(Y_s)^{\otimes i} \right] \right|,$$

then (13f) can be estimated by

$$\begin{aligned} & C_3 |Y_t - \mathcal{V}_{\mathbf{X}_{s,t}} \text{id}(Y_s)| + C_1 |Y_{s,t}|^{\lfloor p \rfloor + 1} + C_2 \omega(s, t)^{\frac{\lfloor p \rfloor + 1}{p}} \\ & \leq C_3 \theta(\omega(s, t)) + C_4 \omega(s, t)^{\frac{\lfloor p \rfloor + 1}{p}}, \end{aligned}$$

for some  $C_4 > 0$  where we have also used that  $Y$  has finite  $p$ -variation controlled by (a scalar multiple of)  $\omega$ . Defining

$$\theta_f(h) := C_3 \theta(h) + C_4 h^{\frac{\lfloor p \rfloor + 1}{p}}$$

finishes the proof. ■

**Remark 3.16** Note that  $\Phi$  is exactly that map that is used in [22] to show that every weakly geometric rough path is also a branched rough path. However, in the proof of Theorem 3.15 we have not used that  $\Phi(\mathbf{X}_{s,t})$  is the image of a weakly geometric rough path. This means that the corresponding result holds for RDEs driven by arbitrary branched rough paths. The corresponding statement for a branched rough path  $\tilde{\mathbf{X}}$  reads as follows with the notation as in Theorem 3.15. We have

$$\left| Y_t - \hat{\Psi}_{V, \text{id}, Y_s}(\tilde{\mathbf{X}}_{s,t}) \right| \leq \theta(\omega(s, t)),$$

if and only for any smooth function  $f$  it holds that

$$\left| f(Y_t) - \hat{\Psi}_{V, f, Y_s}(\tilde{\mathbf{X}}_{s,t}) \right| \leq \theta_f(\omega(s, t)).$$

Hence branched rough paths are only non-geometric in that RDEs driven by them are only well-defined if we fix a connection on the manifolds on which the vector fields are defined. If this manifold is  $\mathbf{R}^n$  there is a canonical choice for this connection. Once a connection (and thus a covariant derivative) is fixed, an RDE driven by a branched rough path satisfies the usual chain rule, since by definition the covariant derivatives of a vector field transform correctly under a change of coordinates. An example of this is an Itô-corrected Stratonovich equation.

**Remark 3.17** From [22, Proposition 3.8] we know that if  $Y$  satisfies (11) it can be lifted to an  $\mathbf{X}$ -controlled rough path  $\mathbf{Y}$  which satisfies the equation  $\mathbf{Y}_t = \mathbf{Y}_s + \int_s^t V(\mathbf{Y}_r) dX_r$  for all  $s, t \in [0, T]$ . The Gubinelli derivatives of  $Y$  are explicitly given by

$$\left( Y_{i_1, \dots, i_k}^{(k)} \right)_s = V_{i_1} \cdots V_{i_k} \text{id}(Y_s)$$

If we replace  $f$  (which corresponds to our  $V$ ) in (3.20) of the proof of [22, Proposition 3.8] by  $Vf$  then the same argument shows that (12) implies that we have

$$f(\mathbf{Y}_t) = f(\mathbf{Y}_s) + \int_s^t Vf(\mathbf{Y}_r) dX_r$$

for all  $s, t \in [0, T]$  and the Gubinelli derivatives of  $f(Y)$  are given by

$$\left( f(Y)_{i_1, \dots, i_k}^{(k)} \right)_s = V_{i_1} \cdots V_{i_k} f(Y_s).$$

It follows from [20, Lemma 8.4] that this is indeed a controlled rough path. Again, all of this holds for arbitrary branched rough paths.

We can now see how Theorem 3.15 guarantees that the solution of an RDE is independent of the choice of coordinates.

**Corollary 3.18** *Let  $\mathbf{X}$  be a weakly geometric rough path and let  $Y$  be a solution of  $dY = V(Y)d\mathbf{X}$ ,  $Y_0 = y_0 \in \mathbf{R}^n$ . Assume that  $Y_t$  lies in some open set  $U \subset \mathbf{R}^n$  for all  $t \in [0, T]$  and  $\varphi : U \rightarrow \varphi(U) \subset \mathbf{R}^n$  is a diffeomorphism. Then  $\tilde{Y} := \varphi(Y)$  solves the rough differential equation  $d\tilde{Y} = \tilde{V}(\tilde{Y})d\mathbf{X}$ ,  $\tilde{Y}_0 = \varphi(y_0)$ , where  $\tilde{V}_a(\tilde{y}) := \varphi_* V_a(\varphi^{-1}(\tilde{y}))$  for all  $a \in \mathbf{R}^d$ .*

*Proof.* Note that if we transform the map  $\varphi$  itself into the new coordinates then we have  $\tilde{\varphi}(\tilde{y}) = \varphi(\varphi^{-1}(\tilde{y})) = \tilde{y}$ . Hence  $\varphi$  is transformed into the identity. Therefore we have

$$\tilde{Y}_t - \tilde{\mathcal{V}}_{\mathbf{X}_{s,t}} \text{id}(\tilde{Y}_s) = \varphi(Y_t) - \tilde{\mathcal{V}}_{\mathbf{X}_{s,t}} \tilde{\varphi}(\tilde{Y}_s) = \varphi(Y_t) - \mathcal{V}_{\mathbf{X}_{s,t}} \varphi(Y_s),$$

where the last equality is obtained by simply expressing everything in the original coordinates. The claim follows now directly from Theorem 3.15 ■

Another application of Theorem 3.15 is the following alternative characterisation of Definition 2.4.

**Corollary 3.19** *Let  $\mathbf{X}$  be a weakly geometric rough path controlled by  $\omega$ . Then a continuous path  $Y : [0, T] \rightarrow \mathbf{R}^n$  is a solution of  $dY = V(Y)d\mathbf{X}$  in the sense of Definition 2.4 if and only if for all compactly supported test functions  $\varphi \in C_c^\infty(\mathbf{R}^n)$  we have*

$$|\varphi(Y_t) - \mathcal{V}_{\mathbf{X}_{s,t}}\varphi(Y_s)| \leq \theta_\varphi(\omega(s, t)) \quad (14)$$

for some function  $\theta_\varphi : \mathbf{R} \rightarrow \mathbf{R}$  with  $\theta_\varphi(h)/h \rightarrow 0$  for  $h \rightarrow 0$  and all  $0 \leq s < t \leq T$ .

*Proof.* It follows directly from Theorem 3.15 that (14) holds whenever  $Y$  is a solution. For the converse statement, let  $K \subset \mathbf{R}^n$  be a compact set which contains  $Y_t$  for all  $t \in [0, T]$ . Let  $\varphi \in C_c^\infty(\mathbf{R}^n)$  be a map which satisfies  $\varphi(x) = \pi_i(x)$  for all  $x \in K$ , where  $\pi_i$  is the projection to the  $i$ -th coordinate for  $i \in \{1, \dots, n\}$ . Then we have

$$\varphi(Y_t) - \mathcal{V}_{\mathbf{X}_{s,t}}\varphi(Y_s) = \pi_i(Y_t) - \mathcal{V}_{\mathbf{X}_{s,t}}\pi_i(Y_s),$$

whence the claim follows since this holds for every  $i \in \{1, \dots, n\}$ . ■

## 4 Rough differential equations on manifolds

Several frameworks which bring together rough paths and manifolds have been developed, such as [2, 6, 10, 14]. They all have in common that they establish notions of rough paths on manifolds. This contrasts with our aim to consider the solution of an RDE which is driven by a vector-space valued rough path along vector fields on a manifold. In Lyons' original approach, which introduces a generalised Picard iteration, the solution of an RDE is necessarily given by a rough path. This is not the case if we use Definition 2.4. Hence we do not need the notion of a rough path on a manifold because the solution will be an ordinary manifold-valued path.

The content of this section is partly inspired by [6]. However, they work entirely on submanifolds of Euclidean space. Even though we will obtain some key results through embedding arguments, we will work on abstract manifolds most of the time.

### 4.1 Definition, local existence and uniqueness

Let  $\mathcal{M}$  be an  $n$ -dimensional smooth manifold and let  $V : \mathbf{R}^d \rightarrow \Gamma(T\mathcal{M})$  be a linear map into the space of smooth vector fields on  $\mathcal{M}$ . Furthermore, let  $\mathbf{X}$  be a weakly geometric  $p$ -rough path on  $\mathbf{R}^d$  for some  $p \geq 1$  which is controlled by a control  $\omega$ . Consider the rough differential equation

$$dY = V(Y)d\mathbf{X}, \quad Y_0 = y_0 \in \mathcal{M}. \quad (15)$$

**Definition 4.1** A continuous path  $Y : [0, T] \rightarrow \mathcal{M}$  is called a *solution* of (15) if for every  $\varphi \in C_c^\infty(\mathcal{M})$  there exists a function  $\theta_\varphi : \mathbf{R} \rightarrow \mathbf{R}$  with  $\theta_\varphi(h)/h \rightarrow 0$  for  $h \rightarrow 0$  such that we have

$$|\varphi(Y_t) - \mathcal{V}_{\mathbf{X}_{s,t}}\varphi(Y_s)| \leq \theta_\varphi(\omega(s, t)) \quad (16)$$

for every  $0 \leq s < t \leq T$ .

We can now use Theorem 2.6 in order to obtain local existence and uniqueness for RDEs on manifolds.

**Theorem 4.2** *The rough differential equation (15) has a unique solution for every initial value  $y_0 \in \mathcal{M}$ . Furthermore this solution will either exist on the whole interval  $[0, T]$  or there will be a time  $\tau \in [0, T]$  such that the solution exists on every interval  $[0, \tilde{T}]$  with  $\tilde{T} < \tau$  and  $\{Y_t : 0 \leq t < \tau\}$  is not relatively compact.*

*Proof.* A similar argument to ours has e.g. been given by Ikeda and Watanabe [26] for SDEs. Fix a locally finite coordinate covering  $(U_i, \psi_i)_{i \in \mathcal{I}}$  of  $\mathcal{M}$  and assume that all coordinate neighbourhoods are relatively compact and that the closure of any coordinate neighbourhood  $U_i$  is contained in another coordinate neighbourhood  $\tilde{U}_i$ .

Now fix  $i \in \mathcal{I}$  such that  $y_0 \in U_i$  and represent the vector fields  $V$  in the local coordinates of  $U_i$  and denote these representations by  $\tilde{V}$ . Since  $U_i$  and consequently  $\psi_i(U_i)$  are relatively compact, we can extend the vector fields  $\tilde{V}$  to smooth and bounded vector fields on  $\mathbf{R}^n$ . Now consider the RDE

$$d\tilde{Y} = \tilde{V}(\tilde{Y})d\mathbf{X}, \quad \tilde{Y}_0 = \psi_i(y_0) \in \psi_i(U)$$

on  $\mathbf{R}^n$ . It follows from Theorem 2.6 that it has a unique solution  $\tilde{Y}$  on the whole interval  $[0, T]$ .

Define  $\tau_1^i := \inf\{t : \tilde{Y}_t \notin \psi_i(U_i)\} \wedge T$ . If  $U_j$  is another neighbourhood of  $y_0$ , we obtain a solution  $\hat{Y}$  in  $\psi_j(U_j)$ . However, it follows from Corollary 3.18 that both solutions coincide on  $\mathcal{M}$  up to time  $\tau_1^i \wedge \tau_1^j$ . Hence the solution is independent of the chosen coordinates.

Now define  $\tau_1 := \max\{\tau_1^i : i \in \mathcal{I}, y_0 \in U_i\}$ . If  $\tau_1 = T$ , then there exists a neighbourhood of  $y_0$  which contains the whole solution. (Recall that we have assumed that the closure of every  $U_i$  is contained in another coordinate neighbourhood.) Otherwise we proceed iteratively and consider the solution which starts at  $Y_{\tau_1}$  at time  $\tau_1$ . This gives us a time  $\tau_2$  and so forth. Now there are two cases. Either there exists  $k \in \mathbf{N}$  with  $\tau_k = T$ , in which case the solution exists globally, i.e. on the whole interval  $[0, T]$ . Or we have  $\tau_k < T$  for all  $k \in \mathbf{N}$ , in which case the solution exists up to every time  $\tilde{T} < \tau := \lim_{k \rightarrow \infty} \tau_k$  but not on the compact interval  $[0, \tau]$ . In this case  $\{Y_t : 0 \leq t < \tau\}$  cannot be relatively compact because it has, by construction, a locally finite cover consisting of infinitely many open sets.

To finish the proof, let us verify that the solution satisfies Definition 4.1. Fix a compact interval  $[0, \tilde{T}]$  on which the solution exists. By the above construction this interval splits into subintervals  $[0, \tau_1], \dots, [\tau_k - 1, T]$ , such that there exist coordinate neighbourhoods  $U_1, \dots, U_k$  with  $Y_t \in U_i$  for  $t \in [\tau_{i-1}, \tau_i]$ . It follows from Corollary 3.19 that (16) holds for every of the subintervals and hence it holds for the whole interval  $[0, T]$ . ■

**Corollary 4.3** *Let  $\mathcal{M}$  be a submanifold of  $\mathbf{R}^k$  and let  $V : \mathbf{R}^d \rightarrow \Gamma(\mathbf{R}^k)$  be a collection of smooth vector fields in  $\mathbf{R}^k$ . Assume that  $Y$  is a solution of  $dY = V(Y)d\mathbf{X}$ ,  $Y_0 = y_0 \in \mathcal{M}$  on the interval  $[0, T]$ . Then we have  $Y_t \in \mathcal{M}$  for all  $t \in [0, T]$ . Furthermore  $Y$  coincides with the intrinsic solution whose existence is guaranteed by Theorem 4.2.*



*Proof.* Denote by  $\tilde{Y}$  the intrinsic solution of the equation whose existence follows from Theorem 4.2. Since functions in  $C_c^\infty(\mathcal{M})$  are restrictions of functions in  $C_c^\infty(\mathbf{R}^k)$  to  $\mathcal{M}$ , it follows from Corollary 3.19 that  $\tilde{Y}$  is also a solution of the same RDE in  $\mathbf{R}^k$ . The uniqueness of the solution implies  $Y = \tilde{Y}$ .  $\blacksquare$

## 4.2 Global existence

Global existence of solutions is a more complicated issue. It requires one to control the growth of the vector fields and their derivatives. An easy way to achieve this is the following assumption.

**Condition 4.4** *Let  $\mathcal{M}$  be a manifold and let  $V$  be a smooth vector field on  $\mathcal{M}$ . We assume that there exists a closed embedding of  $\mathcal{M}$  into  $\mathbf{R}^k$  and a vector field  $\tilde{V}$  in  $\mathbf{R}^k$  such that the pushforward of  $V$  under the embedding map coincides with  $\tilde{V}$  and  $\tilde{V}$  is bounded and has bounded derivatives of all order.*

**Theorem 4.5** *Let  $\mathcal{M}$  be a manifold and let  $V : \mathbf{R}^d \rightarrow \Gamma(T\mathcal{M})$  be a collection of smooth vector fields which satisfy Condition 4.4. Let furthermore  $\mathbf{X}$  be a weakly geometric  $p$ -rough path in  $\mathbf{R}^d$ . Then the RDE*

$$dY = V(Y)d\mathbf{X}, \quad Y_0 = y_0 \in \mathcal{M}$$

*has a unique global solution.*

**Remark 4.6** It is in fact sufficient for global existence if Condition 4.4 is only satisfied for derivatives up to degree  $[p]$ . However, we will need the boundedness of all derivatives later when we consider Malliavin smoothness.

**Remark 4.7** Condition 4.4 also features in [35] and it is mentioned in [13] as well. Nevertheless it is difficult to see under what circumstance it is satisfied and what restriction it imposes on the vector fields (and perhaps on the manifold itself). However, it seems to be a very challenging problem to find an intrinsic criterion that implies Condition 4.4.

## 5 Densities and smooth measures on manifolds

In order to define what we mean by a smooth measure we introduce the *density bundle* of a smooth manifold  $\mathcal{M}$ . We describe its construction, which can be found in many standard references, see e.g. [31, §3.4.1].

Consider the frame bundle  $GL(\mathcal{M})$  which is a principal bundle with structure group  $GL(n)$  where  $n$  is the dimension of  $\mathcal{M}$ . Now define the product bundle  $E := GL(\mathcal{M}) \times \mathbf{R}$  and consider the equivalence relation  $\sim$  on  $E$  where  $(\mathfrak{f}, r) \in E$  and  $(\mathfrak{f}', r') \in E$  are equivalent if and only if there exists  $A \in GL(n)$  such that

$$(\mathfrak{f}, r) = (\mathfrak{f}' \cdot A, |\det A|^{-1} r')$$

Denote by  $|\Lambda|(T\mathcal{M})$  the set of the equivalence classes. There is a well defined projection  $\pi_\Lambda$  to  $\mathcal{M}$  given by  $\pi_\Lambda([\mathfrak{f}, r]) := \pi(\mathfrak{f})$  if  $[\mathfrak{f}, r]$  is the equivalence class that contains  $(\mathfrak{f}, r)$ .

**Definition 5.1** Let  $\mathcal{M}$  be a smooth manifold. A *density* on  $\mathcal{M}$  is a smooth section in the density bundle, i.e. a smooth map  $h : \mathcal{M} \rightarrow |\Lambda|(T\mathcal{M})$  which satisfies  $\pi_\Lambda \circ h = id$ .

If  $(\mathcal{M}, g)$  is a Riemannian manifold,  $g$  induces the density  $\sqrt{g} : \mathcal{M} \rightarrow |\Lambda|(T\mathcal{M})$  which is defined as

$$\sqrt{g}(x) := \left[ \mathfrak{f}, \sqrt{|\det g_{ij}(x)|} \right],$$

where  $\mathfrak{f}$  is an arbitrary basis of  $T_x\mathcal{M}$  and  $g_{ij}$  is the matrix representation of  $g$  in the basis  $\mathfrak{f}$ . This definition does not depend on the choice of  $\mathfrak{f}$ . An important feature of  $\sqrt{g}$  is that it vanishes nowhere because  $g$  is non-degenerate. Hence for any other given density  $h$  we can always find a smooth function  $\phi : \mathcal{M} \rightarrow \mathbf{R}$  such that  $h$  is given by

$$h(x) = \left[ \mathfrak{f}, \phi(x) \sqrt{|\det g_{ij}(x)|} \right]. \quad (17)$$

Every density on  $\mathcal{M}$  induces a measure on  $\mathcal{M}$  in the following way. Let  $h$  be a density on  $\mathcal{M}$  and let  $f : \mathcal{M} \rightarrow \mathbf{R}_{\geq 0}$  be a non-negative measurable function (measurable with respect to the Borel  $\sigma$ -algebra on  $\mathcal{M}$ ). Fix a partition of unity  $(\rho_i)_{i \in \mathcal{I}}$  which is subordinate to a coordinate cover  $(U_i, \psi_i)_{i \in \mathcal{I}}$ . For  $x \in U_i$  the local coordinate  $\psi_i$  induces a basis  $\mathfrak{f}_x^{\psi_i}$  on  $T_x\mathcal{M}$ . Define  $h^{\psi_i} : U_i \rightarrow \mathbf{R}$  such that  $\left[ \mathfrak{f}_x^{\psi_i}, h^{\psi_i}(x) \right] = h(x)$ . Then the integral of  $f$  with respect to  $h$  is defined to be

$$\int_{\mathcal{M}} f dh := \sum_{i \in \mathcal{I}} \int_{\psi_i(U_i)} (\rho_i f)(\psi_i^{-1}(x)) h^{\psi_i}(x) dx,$$

which is possibly equal to infinity. It follows directly from the change of variable formula that this integral is well defined. As usual we define the integral of an arbitrary measurable function as the difference of the integrals of its positive and its negative part, whenever at least one of these two integrals is finite.

**Definition 5.2** A measure  $\nu$  on  $\mathcal{M}$  is said to be *induced by the smooth density*  $h$  if for any measurable set  $A \subset \mathcal{M}$  we have

$$\nu(A) = \int_{\mathcal{M}} \mathbf{1}_A dh.$$

In this case we also say that  $\nu$  is a *smooth measure*.

**Lemma 5.3** Let  $U \subset \mathbf{R}^n$  be an open set and let  $\mu$  be a finite measure on  $U$ . Assume that for every  $k > 0$  and every tuple  $(d_1, \dots, d_k)$  there exists  $C > 0$  such that for every  $\varphi \in C_c^\infty(U)$  we have

$$\left| \int_U \partial_{d_1} \cdots \partial_{d_k} \varphi d\mu \right| \leq C \|\varphi\|_\infty.$$

Then  $\nu$  has a smooth density on  $U$ .

*Proof.* This proof is essentially a local version of the proof from [21]. See [1, Theorem 4.12] for the respective embedding theorems for subsets of  $\mathbf{R}^n$ . For  $s > n/2$  we can embed the Sobolev space  $H_c^s(U)$  into  $C_b(U)$ , i.e. there exists  $C'$  such that

$$\left| \int_U \partial_{d_1} \cdots \partial_{d_k} \varphi d\nu \right| \leq C' \|\varphi\|_{H_c^s(U)}.$$

holds for all  $\varphi \in C_c^\infty(U)$ . Since  $C_c^\infty(U)$  is dense in  $H_c^s(U)$  (by definition) this implies that every distributional derivative of  $\nu$  lies in the dual space  $H^{-s}(U)$  of  $H_c^s(U)$ . Therefore  $\nu$  lies in  $H^l(U)$  for arbitrarily large  $l \in \mathbf{R}$ . This finishes the proof because for  $l > k + n/2$  we have that  $H^l(U)$  can be embedded into  $C^k(U)$ , i.e. we have  $\nu \in C^k(U)$  for all  $k$ .  $\blacksquare$

**Proposition 5.4** *Let  $\mathcal{M}$  be a smooth manifold and let  $\nu$  be a measure which satisfies  $\nu(\mathcal{M}) < \infty$ . Assume furthermore that for every  $k \in \mathbf{N}$  and every collection of compactly supported smooth vector fields  $W_1, \dots, W_k$  there exists a constant  $C > 0$  such that for all  $\varphi \in C_c^\infty(\mathcal{M})$  we have*

$$\left| \int_{\mathcal{M}} W_1 \cdots W_k \varphi d\nu \right| \leq C \|\varphi\|_\infty.$$

*Then  $\nu$  is smooth.*

*Proof.* Let  $(U_i, \psi_i)_{i \in \mathcal{I}}$  be a family of relatively compact coordinate neighbourhoods which covers  $\mathcal{M}$  and which is chosen in such a way that for every  $i \in \mathcal{I}$  the closure  $\overline{U_i}$  is contained in another coordinate neighbourhood  $(\hat{U}_i, \hat{\psi}_i)$  with  $\hat{\psi}_i|_{U_i} = \psi_i$ . For any  $(d_1, \dots, d_k) \in \{1, \dots, n\}^k$  with  $k \in \mathbf{N}$  we can define smooth vector fields  $\tilde{W}_1, \dots, \tilde{W}_k$  on  $\mathbf{R}^n$  such that they vanish outside of  $\hat{\psi}_i(\hat{U}_i)$  and  $\tilde{W}_l$  is constantly equal to the  $d_l$ -th basis vector  $e_{d_l}$  on  $\psi_i(U_i)$  for every  $l \in \{1, \dots, k\}$ . Denote by  $W_1, \dots, W_k$  the vector fields on  $\mathcal{M}$  which are supported inside  $\hat{U}_i$  and whose coordinate representations in  $(\hat{U}_i, \hat{\psi}_i)$  are given by  $\tilde{W}_1, \dots, \tilde{W}_k$  respectively. If we define  $\varphi := \tilde{\varphi} \circ \psi_i$  for every  $\tilde{\varphi} \in C_c^\infty(\psi_i(U_i))$ , we obtain

$$\int_{\mathcal{M}} W_1 \cdots W_k \varphi d\nu = \int_{\psi_i(U_i)} \tilde{W}_1 \cdots \tilde{W}_k \tilde{\varphi} d(\nu \circ \psi_i^{-1}) = \int_{\psi_i(U_i)} \partial_{d_1} \cdots \partial_{d_k} \tilde{\varphi} d(\nu \circ \psi_i^{-1})$$

which implies that there exists a  $C > 0$  such that

$$\left| \int_{\psi_i(U_i)} \partial_{d_1} \cdots \partial_{d_k} \tilde{\varphi} d(\nu \circ \psi_i^{-1}) \right| \leq C \|\varphi\|_\infty = C \|\tilde{\varphi}\|_\infty$$

holds for every  $\tilde{\varphi} \in C_c^\infty(\psi_i(U_i))$ . Therefore we can apply the preceding lemma and we see that the measure  $\nu \circ \psi_i^{-1}$  has a smooth density with respect to the Lebesgue measure on  $\psi_i(U_i)$  and we denote this density by  $h^{\psi_i}$ .

It follows now immediately from the change of variable formula that we obtain a well-defined density  $h$  (in the sense of Definition 5.1) on  $\mathcal{M}$  if we define

$$h(x) := \left[ \mathfrak{f}_x^\psi, h^{\psi_i}(x) \right]$$

for every  $x \in U_i$  and all  $i \in \mathcal{I}$ . Then  $h$  does indeed induce the measure  $\nu$  and therefore  $\nu$  is smooth.  $\blacksquare$

## 6 Malliavin calculus on manifolds

Let  $(\mathcal{W}, \mathcal{H}, \mu)$  be an abstract Wiener space, see [17, 30, 32] for details. Let furthermore  $F : \mathcal{W} \rightarrow \mathbf{R}$  be a random variable. The usual way to prove that the law of  $F$  has a smooth density with respect to the Lebesgue measure assumes that  $F$  belongs to the space  $\mathbb{D}^\infty(\mathbf{R})$  of *Malliavin smooth* functions. There are various ways to define those spaces, a good reference is [34]. The crucial property of random variables in  $\mathbb{D}^\infty(\mathbf{R})$  is that they have moments of every order, that they are infinitely differentiable and that all derivatives also have moments of every order. If we replace  $\mathbf{R}$  by a manifold  $\mathcal{M}$  this becomes problematic because it is not clear what it means for an  $\mathcal{M}$  valued random variable to have moments.

In [35] Taniguchi introduces the following space of manifold valued smooth Wiener functionals.

**Definition 6.1** We say that a random variable  $F : \mathcal{W} \rightarrow \mathcal{M}$  belongs to  $\tilde{\mathbb{D}}^\infty(\mathcal{M})$  if for every  $\varphi \in C_c^\infty(\mathcal{M})$  we have  $\varphi \circ F \in \mathbb{D}^\infty(\mathbf{R})$ .

Let us first see how  $\tilde{\mathbb{D}}^\infty(\mathcal{M})$  is related to  $\mathbb{D}^\infty(\mathbf{R})$  in the case  $\mathcal{M} = \mathbf{R}$ . The inclusion  $\mathbb{D}^\infty(\mathbf{R}) \subset \tilde{\mathbb{D}}^\infty(\mathbf{R})$  is obvious. It is also not hard to see that the converse inclusion is false. Consider for instance the case where  $(\mathcal{W}, \mathcal{H}, \mu)$  is the classical Wiener space associated to Brownian motion and  $F$  is given by

$$F(\omega) = \exp(\exp(\omega(1))).$$

Alternatively we can view  $F$  as the random variable

$$F : (\mathbf{R}, \mathcal{N}(0, 1)) \rightarrow \mathbf{R} : x \mapsto \exp(\exp(x)).$$

Neither  $F$  nor any of its derivatives are integrable. Nevertheless  $F$  is contained in  $\tilde{\mathbb{D}}^\infty(\mathbf{R})$ . The important observation is that the integrability condition imposed by the definition of  $\tilde{\mathbb{D}}^\infty(\mathcal{M})$  is still enough to obtain the existence of a density because the regularity of a probability measure is a local property.

For  $F \in \tilde{\mathbb{D}}^\infty(\mathcal{M})$  the *Malliavin derivative* at  $\omega \in \mathcal{W}$  of  $F$  is defined as the map  $\mathcal{D}F(\omega) : \mathcal{H} \rightarrow T_{F(\omega)}\mathcal{M}$  whose action on smooth functions is given by

$$\mathcal{D}F(\omega)[h]\varphi = \mathcal{D}(\varphi \circ F)(\omega)[h]$$

for all  $\varphi \in C_c^\infty(\mathcal{M})$ . We also write  $\mathcal{D}_h F$  instead of  $\mathcal{D}F(\cdot)[h]$ .

In order to define the associated *Malliavin covariance matrix* we have to equip  $\mathcal{M}$  with a Riemannian metric  $g$ . Then  $\mathcal{D}F(\omega) : \mathcal{H} \rightarrow T_{F(\omega)}\mathcal{M}$  becomes a map between Hilbert spaces and therefore we can define its adjoint  $\mathcal{D}F(\omega)^*$  as the map which satisfies

$$\langle \mathcal{D}F(\omega)^* v, h \rangle_{\mathcal{H}} = g(v, \mathcal{D}F(\omega)h)$$

for all  $h \in \mathcal{H}$  and  $v \in T_{F(\omega)}\mathcal{M}$ . The covariance matrix  $C_F$  associated to  $F$  is then defined as

$$C_F(\omega) := \mathcal{D}F(\omega)\mathcal{D}F(\omega)^*,$$

i.e.  $C_F(\omega)$  is an endomorphism of  $T_{F(\omega)}\mathcal{M}$ .

**Theorem 6.2** *Let  $(\mathcal{M}, g)$  be a Riemannian manifold and let  $F \in \tilde{\mathbb{D}}^\infty(\mathcal{M})$ . Denote by*

$$M(\omega) := \inf\{g_{F(\omega)}(v, C_F(\omega)v) : v \in T_{F(\omega)}\mathcal{M}, g_{F(\omega)}(v, v) = 1\}$$

*the least eigenvalue of the Malliavin matrix. If the random variable  $\varphi(F)M^{-1}$  has moments of all order for every  $\varphi \in C_c^\infty(\mathcal{M})$ , then the law of  $F$  on  $\mathcal{M}$  is smooth.*

Before we prove this theorem, let us remark that even though the covariance matrix  $C_F$  and the random variable  $M$  depend on the metric  $g$ , it can be chosen arbitrarily. The reason for this is that the multiplication by  $\varphi(F)$  makes the behaviour of  $g$  outside of compact sets irrelevant.

In order to prove Theorem 6.2 we need the following lemma.

**Lemma 6.3** *Let  $(\mathcal{M}, g)$  be a Riemannian manifold and let  $F \in \tilde{\mathbb{D}}^\infty(\mathcal{M})$ . Denote by*

$$M(\omega) := \inf\{g_{F(\omega)}(v, C_F(\omega)v) : v \in T_{F(\omega)}\mathcal{M}, g_{F(\omega)}(v, v) = 1\}$$

*the smallest eigenvalue of the Malliavin matrix and assume that the random variable  $\varphi(F)M^{-1}$  has moments of all order for every  $\varphi \in C_c^\infty(\mathcal{M})$ . Let furthermore  $W \in \Gamma_c(T\mathcal{M})$  be a compactly supported smooth vector field. Then the map  $W^F : \mathcal{W} \rightarrow \mathcal{H}$  given by*

$$W^F(\omega) := \mathcal{D}F(\omega)^* C_F(\omega)^{-1} W(F(\omega)) \tag{18}$$

*is an element of  $\mathbb{D}^\infty(\mathcal{H})$ .*

*Proof.* We may assume that the support of  $W$  is contained in a relatively compact coordinate neighbourhood  $(U, \psi)$ . Choose a smooth function  $\varphi : \mathcal{M} \rightarrow \mathbf{R}$  which satisfies  $\varphi|_{\text{supp}(W)} = 1$  and  $\text{supp}(\varphi) \subset U$  and define  $\tilde{F}(\omega) := \psi(F(\omega))\varphi(F(\omega))$  for  $F(\omega) \in U$  and  $\tilde{F}(\omega) = 0$  otherwise. It follows immediately that  $\tilde{F} \in \mathbb{D}^\infty(\mathbf{R}^n)$ . Finally denote by  $\widetilde{W}$  the representation of  $W$  in the coordinate  $\psi$ .

If we can show that

$$W^F(\omega) = \mathcal{D}\tilde{F}(\omega)^* C_{\tilde{F}}(\omega)^{-1} \widetilde{W}(\tilde{F}(\omega)) \tag{19}$$

holds for almost every  $\omega \in \mathcal{W}$ , then we are done because then the claim follows by well known results [32] from  $\mathbf{R}^n$ -valued Malliavin calculus.

Let us therefore prove (19). If  $F(\omega) \notin \text{supp}(W)$ , then both sides of (19) are equal to 0. Hence we assume  $F(\omega) \in \text{supp}(W)$ . By the definition of  $\varphi$  this implies that we have  $\mathcal{D}\tilde{F}(\omega) = d\psi(F(\omega))\mathcal{D}F(\omega)$  and  $\tilde{W}(\tilde{F}(\omega)) = d\psi(F(\omega))W(F(\omega))$ . Therefore we obtain

$$\begin{aligned} & \mathcal{D}\tilde{F}(\omega)^* C_{\tilde{F}}^{-1}(\omega) \tilde{W}(\tilde{F}(\omega)) \\ &= \mathcal{D}F(\omega)^* d\psi(F(\omega))^* (d\psi(F(\omega))\mathcal{D}F(\omega)\mathcal{D}F(\omega)^* d\psi(F(\omega))^*)^{-1} d\psi(F(\omega))W(F(\omega)) \\ &= \mathcal{D}F(\omega)^* (\mathcal{D}F(\omega)\mathcal{D}F(\omega)^*)^{-1} W(F(\omega)) = W^F(\omega). \end{aligned}$$

■

**Remark 6.4** The linear map  $\mathcal{D}F(\omega)^* (\mathcal{D}F(\omega)\mathcal{D}F(\omega)^*)^{-1}$  is also known as the Moore-Penrose pseudoinverse of  $\mathcal{D}F(\omega)$ . This means that  $W^F(\omega)$  can also be characterised by

$$W^F(\omega) = \text{argmin}\{\|h\|_{\mathcal{H}} : \mathcal{D}F(\omega)h = W(F(\omega))\}.$$

In particular,  $W^F$  does not depend on the choice of the Riemannian metric  $g$ , which can also be verified by an easy direct computation. For more details on generalised inverses, see [5].

*Theorem 6.2.* The proof follows [13, Section 8]. Our aim is to apply Proposition 5.4. Therefore we have to show that for every  $k \in \mathbf{N}$  and every collection of compactly supported smooth vector fields  $W_1, \dots, W_k$  there exists  $C > 0$  such that for every  $\varphi \in C_c^\infty(\mathcal{M})$  the estimate

$$\mathbb{E}[|W_1 \cdots W_k \varphi(F)|] \leq C \|\varphi\|_\infty$$

holds. To this end, we define  $W^F : \mathcal{W} \rightarrow \mathcal{H}$  as in Lemma 6.3, i.e. we set

$$W^F(\omega) := (\mathcal{D}F(\omega))^* C_F^{-1} W(F(\omega)).$$

To every test function  $\varphi \in C_c^\infty(\mathcal{M})$  we associate its gradient  $\nabla\varphi \in \Gamma(T\mathcal{M})$  which is given as the canonical dual of the derivative  $d\varphi \in \Gamma(T^*\mathcal{M})$  with respect to  $g$ . Thus we can rewrite the Malliavin derivative of  $\varphi \circ F$  as

$$\mathcal{D}(\varphi \circ F)(\omega)[h] = g_{F(\omega)}(\mathcal{D}F(\omega)h, \nabla\varphi(F(\omega))) = \langle h, (\mathcal{D}F(\omega))^* \nabla\varphi(F(\omega)) \rangle_{\mathcal{H}}.$$

In particular we can plug in  $h = W^F(\omega)$  and the result can be rewritten as

$$\begin{aligned} & \langle W^F(\omega), (\mathcal{D}F(\omega))^* \nabla\varphi(F(\omega)) \rangle_{\mathcal{H}} \\ &= \langle (\mathcal{D}F(\omega))^* (\mathcal{D}F(\omega)(\mathcal{D}F(\omega)^*)^{-1} W(F(\omega))), (\mathcal{D}F(\omega))^* \nabla\varphi(F(\omega)) \rangle_{\mathcal{H}} \\ &= g_{F(\omega)}(W(F(\omega)), \nabla\varphi(F(\omega))) \\ &= (W\varphi)(F(\omega)). \end{aligned}$$

Summarising, we have just shown

$$\mathcal{D}(\varphi \circ F)(\omega)[W^F(\omega)] = (W\varphi)(F(\omega)). \quad (20)$$

For every  $f \in \mathbb{D}^\infty(\mathbf{R})$  we define

$$W^F f(\omega) := \mathcal{D}_{W^F(\omega)} f(\omega).$$

Note that  $W^F$  maps  $\mathbb{D}^\infty(\mathbf{R})$  to itself. In this notation (20) becomes

$$W^F(\varphi \circ F)(\omega) = (W\varphi)(F(\omega)). \quad (21)$$

Now we introduce the  $L^2$ -adjoint  $W^{F\dagger}$  of  $W^F$ . In order to show that  $W^{F\dagger}$  is well-defined on  $\mathbb{D}^\infty(\mathbf{R})$  we observe that for two functions  $f, g \in \mathbb{D}^\infty(\mathbf{R})$  we have

$$\mathbf{E}[W^F f \cdot g + f \cdot W^F g] = \mathbf{E}[W^F(fg)] = \mathbf{E}[\langle \nabla_{\mathcal{D}}(fg), W^F \rangle_{\mathcal{H}}] = \mathbf{E}[fg \cdot \delta W^F] = \mathbf{E}[f \cdot g \delta W^F],$$

where  $\nabla_{\mathcal{D}} f \in \mathcal{H}$  with  $f \in \mathbb{D}^\infty(\mathbf{R})$  is defined by

$$\langle \nabla_{\mathcal{D}} f, h \rangle_{\mathcal{H}} = \mathcal{D}_h f$$

for all  $h \in \mathcal{H}$  and  $\delta$  is its  $L^2$ -adjoint. This can be rewritten as

$$\mathbf{E}[W^F f \cdot g] = \mathbf{E}[f \cdot (g \delta W^F - W^F g)].$$

This last expression shows that

$$(W^{F\dagger} g) = g \delta W^F - W^F g,$$

whence  $W^{F\dagger} g$  belongs to  $\mathbb{D}^\infty(\mathbf{R})$  whenever  $g$  does, because it follows from [32, Prop 1.3.2] that  $\delta W^F$  belongs to  $\mathbb{D}^\infty(\mathbf{R})$ .

It follows from (21) with an easy induction that for vector fields  $W_1, \dots, W_k \in \Gamma(T\mathcal{M})$  we have

$$(W_1 \cdots W_k \varphi)(F) = W_1^F \cdots W_k^F(\varphi \circ F).$$

Therefore we can take expectations and we finally obtain

$$\begin{aligned} |\mathbf{E}[W_1 \cdots W_k \varphi](F)| &= |\mathbf{E}[W_1^F \cdots W_k^F(\varphi \circ F)]| = \left| \mathbf{E} \left[ \varphi(F) W_k^{F\dagger} \cdots W_1^{F\dagger} 1 \right] \right| \\ &\leq \left\| W_k^{F\dagger} \cdots W_1^{F\dagger} 1 \right\|_{L^1} \|\varphi\|_\infty, \end{aligned}$$

which finishes the proof. ■

## 7 Regularity of laws on manifolds under Hörmander's condition

In this chapter we formulate and prove our main result, namely the smoothness of the law of an RDE solution at a fixed time. This means that we will essentially adapt the proof that is given in [9] to the setting that we have introduced in the preceding chapters. We will use an embedding argument which means that we can use almost all of the results from [9] without any change.

Throughout this chapter we assume that  $\mathbf{X}$  is the rough path lift of a  $d$ -dimensional Gaussian process which satisfies the following four conditions from [9].

**Condition 7.1** *Let  $(X_t)_{t \in [0, T]}$  be a Gaussian process in  $\mathbf{R}^d$  with i.i.d. components and suppose that the covariance function has finite Hölder-controlled  $\rho$ -variation for some  $\rho \in [1, 2)$ . We assume that there exist  $p, q \geq 1$  satisfying  $1/p + 1/q > 1$  such that*

- (i)  *$X$  has a natural lift to a geometric  $p$ -rough path and*
- (ii) *the Cameron-Martin space  $\mathcal{H}$  associated with  $X$  can be continuously embedded into the space  $C^{q-\text{var}}([0, T], \mathbf{R}^d)$ .*

**Condition 7.2** *Let  $(X_t)_{t \in [0, T]}$  be a Gaussian process in  $\mathbf{R}^d$  with i.i.d. components and suppose that the covariance function has finite Hölder-controlled  $\rho$ -variation for some  $\rho \in [1, 2)$ . We assume that there exists  $\alpha > 0$  such that*

$$\inf_{0 \leq s < t \leq T} \frac{1}{(t-s)^\alpha} \text{Var}(X_{s,t} \mid \mathcal{F}_{0,s} \vee \mathcal{F}_{t,T}) > 0,$$

where  $\mathcal{F}_{a,b} := \sigma(X_{s,t} : a \leq s \leq t \leq b)$ .

**Condition 7.3** *Let  $(X_t)_{t \in [0, T]}$  be a continuous, real-valued Gaussian process. We assume that for every  $[u, v] \subset [s, t] \subset [0, S] \subset [0, T]$  we have*

$$\text{Cov}(X_{s,t} X_{u,v} \mid \mathcal{F}_{0,s} \vee \mathcal{F}_{t,S}) \geq 0.$$

**Condition 7.4** *Let  $(X_t)_{t \in [0, T]}$  be a continuous, real-valued Gaussian process. We assume that for every  $0 < S < T$  and every partition  $D = \{0 = t_0 < \dots < t_n = S\}$  the matrix*

$$Q_{ij}^D := \mathbf{E}[X_{t_{i-1}, t_i} X_{t_{j-1}, t_j}]$$

*is diagonally dominant.*

Furthermore we assume that  $\mathcal{M}$  is a smooth manifold and  $V = V_1, \dots, V_d$  is a collection of smooth vector fields which satisfy Condition 4.4. This means that the rough differential equation (15) has a unique solution for every realisation of  $\mathbf{X}$  and this solution exists globally almost surely. Hence for every time  $t > 0$  we can consider the well-defined  $\mathcal{M}$ -valued random variable  $Y_t$ . We want to apply the theory of the previous chapter in order to show that this random variable has a smooth law.



The invertibility of the Malliavin matrix is closely related to the Lie brackets of the vector fields  $V_1, \dots, V_d$ . Therefore we define

$$\begin{aligned}\mathcal{V}_1 &:= \{V_1, \dots, V_d\} \\ \mathcal{V}_n &:= \{[V_i, W] : i \in \{1, \dots, d\}, W \in \mathcal{V}_{n-1}\}.\end{aligned}$$

Since any subspace of  $T_{y_0}\mathcal{M}$  can be at most  $n$ -dimensional, there exists an integer  $N$  such that we have

$$Z_V(y_0) := \text{span} \left\{ W(y_0) : W \in \bigcup_{i=1}^{\infty} \mathcal{V}_i \right\} = \text{span} \left\{ W(y_0) : W \in \bigcup_{i=1}^N \mathcal{V}_i \right\} \subset T_{y_0}\mathcal{M}.$$

We say that  $V$  satisfies *Hörmander's condition* if  $Z_V(y_0) = T_{y_0}\mathcal{M}$  holds. Note that this is only a statement about the vector fields at the initial value and they could still be more degenerate elsewhere. In contrast, Taniguchi [35] works with a global Hörmander condition, i.e.  $Z_V(y) = T_y\mathcal{M}$  for every  $y \in \mathcal{M}$ .

**Theorem 7.5** *Let  $\mathbf{X}$  be a Gaussian rough path which satisfies the Conditions 7.1 to 7.4. Let  $\mathcal{M}$  be a smooth manifold and let  $V = (V_1, \dots, V_d)$  be a collection of smooth vector fields which satisfy Condition 4.4. Assume furthermore that  $V$  satisfies Hörmander's condition. Then for any  $t > 0$  the random variable  $Y_t : \mathcal{W} \rightarrow \mathcal{M}$  which is given as the solution of (15) at time  $t$  has a smooth law.*

We want to prove this theorem by an application of Theorem 6.2. Therefore we first have to show that  $\mathcal{M}$  is isomorphic to a submanifold  $\tilde{\mathcal{M}}$  of  $\mathbf{R}^k$  in such a way that the corresponding map  $\tilde{Y}_t : \mathcal{W} \rightarrow \tilde{\mathcal{M}}$  is in  $\mathbb{D}^\infty(\mathbf{R}^k)$ . This is however an immediate consequence of Condition 4.4 in connection with [9, Proposition 7.5].

It remains to verify that the inverse Malliavin matrix satisfies the necessary integrability condition. We do this by slightly changing the the proof of Theorem 3.5 in the last section of [9]. The following lemma follows immediately from the results in [11].

**Lemma 7.6** *The inverse of the smallest eigenvalue of the Malliavin matrix  $C_{Y_t}$  has moments of all order if and only if the same is true for the inverse of the smallest eigenvalue of the reduced Malliavin matrix*

$$\bar{C}_{Y_t} := ((\Phi_t^{-1})_* \mathcal{D}Y_t) ((\Phi_t^{-1})_* \mathcal{D}Y_t)^* \in T_{y_0}^*\mathcal{M} \otimes T_{y_0}\mathcal{M}.$$

It is easier to work with  $\bar{C}_{Y_t}$  because we only assume that Hörmander's condition holds at the starting point and  $\bar{C}_{Y_t}$  is an endomorphism on  $T_{y_0}\mathcal{M}$  whereas  $C_{Y_t}$  is an endomorphism on  $T_{Y_t}\mathcal{M}$ . The lemma shows that working with  $\bar{C}_{Y_t}$  does not change the result.

Observe that the proof of Theorem 3.5 in the last section of [9] contains already precisely the result that we need, namely the following.

**Theorem 7.7** *Let  $V = (V_1, \dots, V_d)$  be a collection of smooth and bounded vector fields on  $\mathbf{R}^k$  which have bounded derivatives of every order and let  $\mathbf{X}$  be a Gaussian rough path in  $\mathbf{R}^d$  which satisfies Conditions 7.1 to 7.4. Let  $\bar{C}_{Y_t}$  be the reduced Malliavin matrix of the solution  $Y_t$  of the equation  $dY = V(Y)d\mathbf{X}$ . Then the random variable*

$$\left( \inf \left\{ |v^t \bar{C}_{Y_t} v| : v \in Z_V(y_0), |v| = 1 \right\} \right)^{-1}$$

*has moments of every order for every  $t > 0$ .*

In the original formulation in [9] the authors assume at this point that Hörmander's condition holds and therefore they take the infimum over all  $v \in \mathbf{R}^k$  instead of  $v \in Z_V(y_0)$ . In our setting, where  $\mathbf{R}^k$  only serves as the ambient space for an embedding of  $\mathcal{M}$ , Hörmanders condition does in general not imply that  $Z_{\tilde{V}}(\tilde{y}_0) = \mathbf{R}^k$ .

*Theorem 7.5.* Let  $Y_t$  be the solution of (15) and let  $\tilde{Y}_t$  be the image of  $Y_t$  under some embedding  $\Psi$  as in Condition 4.4. Then we can identify the reduced Malliavin matrix  $\bar{C}_{Y_t}$  with the restriction of  $\bar{C}_{\tilde{Y}_t}$  to the subspace  $T_{\Psi(y_0)}\tilde{\mathcal{M}} \subset \mathbf{R}^k$ . Then Theorem 7.7 together with Hörmander's condition  $Z_V(y_0) = T_{y_0}\mathcal{M}$  implies that the inverse of the smallest eigenvalue of  $\bar{C}_{Y_t}$  has finite moments of every order. In combination with the Malliavin smoothness of  $Y_t$  the conclusion of Theorem 7.5 follows now from Theorem 6.2.  $\blacksquare$

## References

- [1] Robert A. Adams and John J. F. Fournier, *Sobolev spaces*, Academic Press, 2003.
- [2] Ismaël Bailleul, *Rough integrators on Banach manifolds*, arXiv:1403.3285, 2014.
- [3] Fabrice Baudoin and Martin Hairer, *A version of Hörmander's theorem for the fractional Brownian motion*, Probab. Theory Related Fields **139** (2007), no. 3, 373–395.
- [4] Fabrice Baudoin, Eulalia Nualart, Cheng Ouyang, and Samy Tindel, *On probability laws of solutions to differential systems driven by a fractional Brownian motion*, Ann. Probab., to appear.
- [5] Adi Ben-Israel and Thomas N.E. Greville, *Generalized inverses*, Springer-Verlag New York, 2003.
- [6] Thomas Cass, Bruce K. Driver, and Christian Litterer, *Constrained rough paths*, Proceedings of the London Mathematical Society **111** (2015), no. 6, 1471–1518.
- [7] Thomas Cass, Peter Friz, and Nicolas Victoir, *Non-degeneracy of Wiener functionals arising from rough differential equations*, Transactions of the American Mathematical Society **361** (2009), no. 6, 3359–3359.

- [8] Thomas Cass and Peter K Friz, *Densities for rough differential equations under Hörmander's condition*, Ann. of Math. **171** (2010), no. 2, 2115–2141.
- [9] Thomas Cass, Martin Hairer, Christian Litterer, and Samy Tindel, *Smoothness of the density for solutions to Gaussian rough differential equations*, Ann. Probab. **43** (2015), no. 1, 188–239.
- [10] Thomas Cass, Christian Litterer, and Terry Lyons, *Rough paths on manifolds*, New trends in stochastic analysis and related topics, Interdiscip. Math. Sci., vol. 12, World Sci. Publ., Hackensack, NJ, 2012, pp. 33–88.
- [11] ———, *Integrability and tail estimates for Gaussian rough differential equations*, Ann. Probab. **41** (2013), no. 4, 3026–3050.
- [12] Alexander M Davie, *Differential equations driven by rough paths: An approach via discrete approximation*, Appl. Math. Res. Express (2008).
- [13] Bruce K Driver, *Curved Wiener space analysis*, Real and Stochastic Analysis (M M Rao, ed.), Trends in Mathematics, Birkhäuser Boston, 2004, pp. 43–198.
- [14] Bruce K Driver and Jeremy Semko, *Controlled Rough Paths on Manifolds I*, arXiv:1504.03308, 2015.
- [15] Peter K Friz and Martin Hairer, *A course on rough paths with an introduction to regularity structures*, Universitext, Springer, 2014.
- [16] Peter K Friz and Nicolas B Victoir, *Multidimensional stochastic processes as rough paths. theory and applications*, Cambridge University Press, 2010.
- [17] Leonard Gross, *Abstract Wiener spaces*, Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. II: Contributions to Probability Theory, Part 1, Univ. California Press, Berkeley, Calif., 1967, pp. 31–42.
- [18] Robert Grossman and Richard G Larson, *Hopf-algebraic structure of families of trees*, J. Algebra **126** (1989), no. 1, 184–210.
- [19] Massimiliano Gubinelli, *Controlling rough paths*, Journal of Functional Analysis **216** (2004), no. 1, 86–140.
- [20] ———, *Ramification of rough paths*, Journal of Differential Equations **248** (2010), no. 4, 693–721.
- [21] Martin Hairer, *On Malliavin's proof of Hörmander's theorem*, Bull. Sci. Math. **135** (2001), no. 6–7, 650–666.

- [22] Martin Hairer and David Kelly, *Geometric versus non-geometric rough paths*, Annales de l'institut Henri Poincaré (B) Probability and Statistics **51** (2015), no. 1, 207–251.
- [23] Martin Hairer and Natesh S. Pillai, *Regularity of laws and ergodicity of hypoelliptic sdes driven by rough paths.*, Ann. Probab. **41** (2013), no. 4, 2544–2598.
- [24] Michael E. Hoffman, *Combinatorics of rooted trees and Hopf algebras*, Transactions of the American Mathematical Society **355** (2003), no. 09, 3795–3811.
- [25] Elton P. Hsu, *Stochastic analysis on manifolds*, American Mathematical Soc., 2002.
- [26] Nobuyuki Ikeda and Shinzo Watanabe, *Stochastic differential equations and diffusion processes*, second ed., North-Holland Mathematical Library, vol. 24, North-Holland Publishing Company, Amsterdam, 1989.
- [27] Terry J Lyons, *Differential equations driven by rough signals*, Rev. Mat. Iberoamericana **14** (1998), no. 2, 215–310.
- [28] Terry J Lyons, Michael Caruana, and Thierry Lévy, *Differential equations driven by rough paths*, Lecture Notes in Mathematics, vol. 1908, Springer, Berlin, 2007.
- [29] Paul Malliavin, *Stochastic calculus of variation and hypoelliptic operators*, roceedings of the International Symposium on Stochastic Differential Equations (Res. Inst. Math. Sci., Kyoto Univ., Kyoto, 1976) (New York-Chichester-Brisbane), Wiley, 1978, pp. 195–263.
- [30] ———, *Stochastic analysis*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 313, Springer-Verlag, Berlin, 1997.
- [31] Liviu I Nicolaescu, *Lectures on the geometry of manifolds*, World Scientific Publishing Co., Inc., River Edge, NJ, 1996.
- [32] David Nualart, *The Malliavin calculus and related topics*, second ed., Springer-Verlag, Berlin, Heidelberg, 2006.
- [33] Christophe Reutenauer, *Free lie algebras*, Clarendon Press, 1993.
- [34] Hiroshi Sugita, *On a characterization of the Sobolev spaces over an abstract Wiener space*, J. Math. Kyoto Univ. **25** (1985), no. 4, 717–725.
- [35] Setsuo Taniguchi, *Malliavin's stochastic calculus of variations for manifold-valued Wiener functionals and its applications*, Z. Wahrsch. Verw. Gebiete **65** (1983), no. 2, 269–290.